Artificial Intelligence and Marketing: Pitfalls and Opportunities

Arnaud de Bruyn, Vijay Viswanathan, Yean Shan Beh, Jürgen Kai-Uwe Brock, Florian von Wangenheim

VISIT…

LANZAROTE Caliente.com

Version of Record: https://www.sciencedirect.com/science/article/pii/S1094996820300888
Manuscript_f2519d16a5ce83f510182f334056162b

Artificial Intelligence and Marketing: Pitfalls and Opportunities

Arnaud De Bruyn[1]
ESSEC Business School

Vijay Viswanathan
Northwestern University

Yean Shan Beh
The University of Auckland

Jürgen Kai-Uwe Brock
Fujitsu Global

Florian von Wangenheim
ETH Zurich

Abstract – This article discusses the pitfalls and opportunities of AI in marketing through the lenses of knowledge creation and knowledge transfer. First, we discuss the notion of "higher-order learning" that distinguishes AI applications from traditional modeling approaches, and while focusing on recent advances in deep neural networks, we cover its underlying methodologies (multilayer perceptron, convolutional, and recurrent neural networks) and learning paradigms (supervised, unsupervised, and reinforcement learning). Second, we discuss the technological pitfalls and dangers marketing managers need to be aware of when implementing AI in their organizations, including the concepts of badly-defined objective functions, unsafe or unrealistic learning environments, biased AI, explainable AI, and controllable AI. Third, AI will have a deep impact on predictive tasks that can be automated and require little explainability, we predict that AI will fall short of its promises in many marketing domains if we do not solve the challenges of tacit knowledge transfer between AI models and marketing organizations.

[1] Contact author. Email: debruyn@essec.edu. Phone: +33a (0)1 34 43 32 46. Address: ESSEC Business School, Avenue Bernard Hirsch, 95000 Cergy, France.

INTRODUCTION

Despite the profound impact AI is likely to have on a wide array of business functions, it is somehow disheartening to realize the extent to which some managers still have an insufficient understanding of what AI is, what it can do, and what it cannot.

In an executive education setting, a C-level manager ventured to say that "what is amazing with AI is that, with enough data, it can learn anything." Given the relative newness and complexity of the subject matter, such misconceptions were to be expected in the short term but present two classes of challenges nonetheless. First, marketing managers who have been sold on the "magical powers" of AI-driven solutions may underestimate its dangers, limitations, and pitfalls. Second, business managers might misjudge the areas where AI is the most likely to bear fruit –and where it will most likely fail– if adopted by marketing organizations, and therefore misallocate their efforts and resources.

This paper aims at addressing these two challenges and is divided into three sections.

Section §1 is primarily educational and lays the groundwork for the remainder of the paper. We first define AI through the lenses of "higher-order learning." Focusing on the most recent advances in deep learning and artificial neural networks, we then introduce the most common neural network topologies (multilayer perceptrons, convolutional neural networks, and recurrent neural networks), the three learning paradigms in which these networks are trained (supervised learning, unsupervised learning, and reinforcement learning), and how these technologies apply to various marketing contexts.

In §2, we build upon the concept of higher-order learning to discuss related pitfalls and dangers of AI solutions as they pertain to their adoption by marketing organizations. All machine learning enterprises face common challenges such as the need for abundant data, development of

the firm's analytic capabilities (e.g., Kumar et al., 2020), organizational adoption and change management, etc. However, the autonomous generation of higher-order constructs by AI algorithms creates (or aggravates) specific challenges. We focus our attention on these challenges, such as the difficulties to specify valid objective functions, to simulate a safe and realistic learning environment, the risks of –unwillingly– develop and deploy biased, unexplainable, or uncontrollable AI, and the automation paradox.

In §3, we discuss the impending challenges AI applications will face in marketing where tacit knowledge is crucial –and the facile transfer of tacit knowledge among marketing stakeholders a source of competitive advantage–. Recent AI techniques have demonstrated their abilities to learn autonomously from big data and self-generated experience, impervious to human expertise. One could justifiably marvel at these achievements. Nevertheless, we argue that their impermeability to marketing stakeholders' tacit knowledge, while being the source of their early successes, may very well be the cause of their near-term limitations in domains where tacit knowledge is crucial, such as in sales, branding, or relationship marketing.

We conclude by a call for marketing organizations and academic researchers to focus on more efficient tacit knowledge transfer between marketing stakeholders and AI machines.

ARTIFICIAL INTELLIGENCE AND DEEP LEARNING

Defining artificial intelligence

Whether one surveys psychologists, sociologists, biologists, neuroscientists, or philosophers, the term "intelligence" can take more than 70 different definitions (Legg and Hutter, 2007). It is therefore not surprising that the term "artificial intelligence" (AI), while so commonly used,

remains so badly defined, and such a fuzzy concept (Kaplan and Haenlein, 2019). Rather than take a stand, we propose three definitions with varying degrees of inclusiveness.

The most encompassing definition

A widely accepted definition of AI is intelligence demonstrated by machines (Shieber, 2004). In the same vein, Brooks (1991) wrote: "Artificial Intelligence is intended to make computers do things, that when done by people, are described as having indicated intelligence." While valid, this definition only kicks the can down the road, since it assumes a general agreement on the very notion of intelligence itself. While intelligence is most closely associated with learning, planning, and problem-solving (Russel and Norvig, 2016), it may also encompass understanding, self-awareness, emotional knowledge, reasoning, creativity, logic, and critical thinking (Legg and Hutter, 2007).

Based on this widely-accepted definition of AI, and depending on how intelligence is defined or understood, some may argue that we are decades away from achieving AI, while others may consider that a simple regression analysis (which involves learning through the minimization of a loss function) is achieving artificial intelligence already. This quite lax definition has allowed companies to claim they offer AI-powered products and services (a strategy dubbed "AI washing," see Bini, 2018), where most AI researchers would be dubious at best to qualify them as such. Some components of Salesforce's "Einstein AI," for instance, are based on simple regression analyses.

While we can indeed define AI as "intelligence demonstrated by machines," it does not help us clearly define the perimeter of what constitutes AI per se, and what does not, and we need to recognize that this loose definition can lead to confusion, misunderstandings, and abuses.

The most restrictive definition

Several researchers believe that it would be beneficial if one used the term “artificial intelligence” only when referring to “artificial general intelligence” (AGI), that is, the intelligence of a machine that can understand or learn *any* intellectual task that a human being can (Goertzel, 2015; Thorisson et al., 2015).

Today, machines only learn to perform specific, well-defined, and restricted tasks, such as playing chess, recognizing human faces, or predicting the likelihood an online visitor will click on a banner ad. These algorithms are sometimes referred to as “weak AI” or “narrow AI” because they cannot learn anything outside the narrow domain they have been programmed to operate in. However, learning is a cognitive activity, and in theory, any cognitive activity can be learned. Consequently, it might be possible one day to program machines that “learn to learn,” a concept called “strong AI” (Kurzweil, 2005) or “artificial general intelligence.”

Since the general scientific consensus is that we are decades away from AGI, some researchers facetiously define AI as “everything we cannot do yet.”

There is some truth to that statement. Techniques commonly associated with the field of AI in the ‘70s or ‘80s, such as optical character recognition (OCR) or rule-based expert systems, are so ubiquitous today that few would still qualify these as AI techniques[2]. In that sense, as soon as the community masters a technique, understands it well, and adopts it widely, the said technique tends to leave the realm of what is considered AI. It would not surprise the authors if, one day, even deep learning was not considered AI anymore.

In other words, if artificial general intelligence is the ultimate goal, recent progress may only be intermediary, small steps in that direction, and may not deserve the label of AI.

[2] Here, we restrictedly define OCR as the transformation of pixels into characters. The field of OCR is being rejuvenated by modern AI techniques that simultaneously recognize characters *and* comprehend the context.

AI and higher-order learning

On the one hand, AI could be defined extremely loosely, to a point where pretty much *everything* in statistics could be considered AI; On the other hand, one could define AI so restrictively that *nothing* would be (yet).

For this article, we propose a more specific definition of AI as "machines that mimic human intelligence in tasks such as learning, planning, and problem-solving through higher-level, autonomous knowledge creation."

This definition has several advantages: (1) it confines the definition of intelligence to three specific subtasks; (2) it does not claim AI achieves intelligence, but rather that it mimics some of its output –therefore avoiding philosophical debates about whether machines can be intelligent–; and (3) it restricts AI to algorithms that *autonomously* generate new constructs and knowledge structures.

When Anderson et al. (2001) revisited Bloom's taxonomy of educational learning objectives (Bloom et al., 1956), they defined *creation* as the highest-level learning objective. They define it as "Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure."

We will show that what best distinguishes AI algorithms from classical statistical techniques is the notion of knowledge creation in the sense of Anderson's taxonomy. We will later argue that such distinction has profound consequences on the likelihood of adoption of AI technologies in areas of marketing that require or could benefit from knowledge transfer.

Artificial neural networks

A focus on artificial neural networks

It would be misleading to restrict the field of AI to deep learning and artificial neural networks, therefore excluding important subfields such as symbolic AI or expert systems, to only name two. Regardless, as far as business applications in general –and marketing applications in particular– are concerned, one has to reckon that the most recent and impressive progress of AI is related to the booming field of deep learning (e.g., in marketing and services, see Riikkinen et al., 2018; Dimitrieska et al., 2018; Pavaloiu, 2016; Erevelles et al., 2016). A majority of AI applications in the business area refer to the use of deep artificial neural networks to solve complex predictive tasks that were deemed unsolvable less than a decade ago. Among other things, predictive analytics in marketing enables marketers to forecast future marketing actions and its impacting behavior, generate insights to improve leads, acquire new customers, and achieve pricing optimization (Power, 2016; Murray and Wardley, 2014).

Given the current emphasis on artificial neural networks and deep learning, we believe it is important to clarify the *types* of artificial neural networks most commonly used today, namely:

- Multilayer perceptrons (MLP)
- Convolutional neural networks (CNN)
- Recurrent neural networks (RNN)

We will then cover the *learning paradigms* in which these neural networks are trained:

- Supervised learning
- Unsupervised learning
- Reinforcement learning

In this article, we will assume that the reader has a working understanding of what a basic artificial neural network is (for an introduction, see Ramchoun et al., 2016; DaSilva et al., 2017). For overall introductory references to deep learning, we also refer to Goodfellow et al. (2016), Kelleher (2019), and Aggarwal (2018).

Multilayer perceptron (MLP)

A multilayer perceptron (MLP) is the simplest form of feedforward neural network[3], consisting of multiple layers of computational units (or neurons), where each neuron in one layer is directly –and usually fully– connected to the neurons of the subsequent layer, and where the outputs of one layer serve as inputs to the next (e.g., see Chapmann, 2017; Fine, 1999).

The term "deep learning" refers to the number of hidden layers in a neural network. Some modern applications can have dozens or even a hundred or more hidden layers, hence allowing the neural network to learn extremely complex relationships between its inputs and its outputs (Egriogglu et al., 2008).

Multilayer perceptrons are universal prediction models (also dubbed "prediction machines," see Agrawal, Gans, and Goldfarb, 2018) and shine with fixed-length input, that is, when the number of predictors is well defined and known in advance. They are widely used for pattern recognition and classification problems. Applications of MLP include stock price prediction (Di Persio and Honchar, 2016), customer churn prediction (Ismail et al., 2015), credit scoring prediction (Zhao et al., 2015), and evaluating customer loyalty (Ansari and Riasi, 2016).

Convolutional neural networks (CNN)

A convolutional neural network (Khan et al., 2018) is typically a deep-learning, feedforward neural network that contains at least one convolutional layer. Convolutional layers automatically

[3] Note that authors sometimes confound perceptron and feedforward neural networks. While a perceptron is a feedforward neural network, not all feedforward neural networks are fully connected networks like perceptron.

identify patterns in data (usually images), and stacking a sequence of convolutional layers allows the network to identify increasingly complex patterns. For instance, the first layer of a convolutional neural network may learn to identify horizontal, vertical, and diagonal edges in a picture; the second layer might learn to combine these edges to identify eyes, noses, and mouths; and the third layer might combine these patterns to identify visages.

Convolutional neural networks' strength lies in their ability to identify patterns in data regardless of their position. For instance, if a CNN learns to identify the presence of an eye, it can learn to do so regardless of where it is positioned in the picture.

Although CNNs are most widely known for their prowess in image recognition and computer vision, recent applications have demonstrated their value in natural language processing and time-series forecasting as well. For instance, in marketing, if specific *patterns* of behaviors are predictive of a customer churning (e.g., service incident → customer complaint → failure to solve the problem), a CNN might be especially efficient at identifying such patterns for churn predictions.

Recurrent neural networks (RNN)

One of the main limitations of feedforward neural networks (i.e., MLP and CNN) is that the dimensionality of their input needs to be fixed and stable. Take, for instance, a deep neural network tasked with predicting whether a customer is likely to click on an online banner; it requires a clearly-defined and finite set of inputs, or predictors (e.g., age, gender, time of day, inferred online visitor's interests, features summarizing visitor's previous behavior). Likewise, in computer vision, a convolutional neural network will typically be trained on a set of pictures of identical sizes (even if it means padding or resizing the original pictures to similar dimensions) (Wang et al., 2017; Jordan and Mitchell, 2015).

This limitation makes traditional neural networks particularly unsuited for sequence data, where the length of the input data is not fixed or cannot be determined in advance, as it is the case for speech recognition, music generation, machine translation, sentiment classification, or time-series predictions.

As a solution, recurrent neural networks (Mandic and Chambers, 2020) use the output of one timestep as input for the next timestep, hence allowing the neural network to keep useful information in memory for future predictions. Such a structure allows the neural network to learn from (and predict) a sequence of inputs of undefined or varying lengths.

For instance, in natural language processing, the sentence "Tom fell on the floor and broke […]" can be seen as a sequence of 7 consecutive vector inputs. A well-trained neural network would keep in memory that the sentence began by a masculine subject ("Tom"), and therefore that the next word in this particular sentence would be more likely to be "his" than "her." The presence of the words "floor" and "broke" would somehow be kept in memory and used to predict that the next word after "his" is more likely to be "arm" than "car."

One of the most common recurrent neural networks uses Gated Recurrent Units (Cho et al., 2014), or GRU, which learn to identify which information it needs to remember from one timestep to the next. A more complex (and historically earlier) version, the Long Short-Term Memory unit, or LSTM (Gers et al., 1999), also includes a forget gate, specifically designed to learn when prior information becomes irrelevant and should be ignored for future predictions. In word-sequence predictions, for instance, the occurrence of a period indicates the end of a sentence, making a lot of prior remembered information obsolete for future predictions.

Recurrent neural networks are most famous and have proven invaluable in speech recognition, sequence generation (e.g., music, text, voice), text translation, or sentiment analysis (e.g., where a sequence of words of undefined length is used to predict the sentiment valence of

the sentence). However, researchers have demonstrated their applicability recently to predict sequences of behavior in marketing as well. For instance, Sarkar and De Bruyn (2019) have shown that an LSTM network fed with simple, raw behavioral data could outperform a logit model with advanced feature engineering; Valentin and Reutterer (2019) demonstrated the ability of LSTM models to outshine buy-till-you-die models (e.g., Pareto-NBD).

Learning paradigms

Supervised learning

In a supervised learning paradigm, a neural network learns from a set of examples (training data) where both inputs (predictors) and outputs (target variables) are known to the analyst, such as the model learns to minimize a loss function (e.g., entropy).

A major difference between deep neural networks and more traditional techniques such as linear regression, logistic regression, classification and regression trees, or support vector machines, is that deep neural networks can automatically and autonomously identify higher-level constructs in the data.

For instance, a convolutional neural network tasked to recognize visages in images, and fed with a sufficiently-large training data, will automatically learn to identify edges in the images and combine these edges into higher-level constructs such as noses, eyes or mouths, without any human intervention. It might even recognize patterns in the data that the analyst was oblivious to, hence achieving Anderson et al.'s (2001) higher-learning objective ("reorganizing elements into a new pattern or structure").

As the input data travels through the various layers of a deep neural network, the algorithm will autonomously recombine the data into higher-level constructs, effectively identifying and

creating its list of predictive variables ‒a steep departure from classic regression analyses where independent variables have to be determined by the analyst (Zheng and Casari, 2018)‒.

For instance, in the context of customer prediction in a direct marketing context, Sarkar and De Bruyn (2019) showed that an LSTM model automatically identified predictive features such as recency, frequency, and seasonality from raw data, and retained them in memory for future predictions, with no human-crafted feature engineering or prior domain knowledge.

One noteworthy concept is that of *labeled* data. In statistics, the target variable is traditionally quantitative, either binary (e.g., is this customer going to churn?), multinomial (e.g., which brand will she purchase?) or continuous (e.g., what amount will he spend?). Since the data is an observable quantity by nature, no additional data preparation is needed.

Within an AI context, in many supervised learning applications, data often needs to be manually labeled first, and this may be a laborious endeavor. If a company wishes to deploy a deep learning algorithm that automatically predicts whether a customer is angry or not based on his or her most recent email or written comments, the firm first needs to *label* emails from previous customers into angry/not angry categories, and use that data to calibrate a predictive model (Vo et al., 2018). Given how data-hungry deep learning models can be, this may be a serious impediment; and a source of competitive advantage for those companies having access to a large database of potential training (and labeled) data, including knowledge extraction, sentiment detection, and analysis.

Unsupervised learning

Unsupervised learning helps find patterns in data without pre-existing labels. In traditional statistics, the most common unsupervised models include k-means segmentation, hierarchical

clustering, factor analysis, or multidimensional scaling. Unsupervised learning applications rarely make the headlines, but they are an essential component of many successful AI applications.

For instance, in natural language processing, words and tokens were historically encoded as a sparse vector. If the dictionary contained 10,000 words, each word was encoded as a vector of size 10,000, where all but one value was equal to 0. This approach proved unsuitable for complex problems, where dictionaries of 100K words and more were common. Instead, *word embedding* is one of the most popular representations of document vocabulary, where each word is represented as a vector of unlabeled features. These vectors' dimensions are in the hundreds, rather than in the thousands, hence facilitating learning and predictions. Typical of unsupervised learning, these features are learned by parsing a large amount of text, with no prior labeling required. The algorithm identifies relationships between words based on their co-occurrences in a corpus of text, and autonomously creates higher-level constructs that link words together.

Reinforcement learning

Reinforcement learning (Sutton and Barto, 2018) is an area of machine learning where an agent learns to take actions in an environment to maximize rewards and minimize penalties over time. For instance, reinforcement learning has learned to play games such as Chess or Go, and it is an essential component of how autonomous vehicles and robots operate.

Schematically, if A is a set of actions an agent could take, S is the state of the environment, and R is the –long-term, and possibly discounted– reward received by the agent, the problem is to approximate the function $f(S,A) \rightarrow R$ to select the best sequence of actions that leads to the highest stream of (discounted) rewards.

Several interesting challenges arise in reinforcement learning, such as the necessity to balance exploration (to learn f in its entirety) and exploitation (to maximize rewards), or the difficulty

posed by delayed rewards (where the agent may obtain the reward/penalty only at the very last step, such as winning/losing a game of chess, and the algorithm has to learn which actions led to that outcome). The most common reinforcement learning algorithms are TD-learning, SARSA, and Q-learning, to only cite the most common (see Sutton and Barto, 2018 for a review).

Reinforcement learning has been around for decades, but progresses in the field of AI have recently propelled the field to new heights. One of these advances is the ability to deploy deep artificial neural networks, able to take large amounts of information as inputs –i.e., the state of the environment *S*, that could now potentially be described by thousands or tens of thousands of indicators–, and learn to autonomously discover complex relationships between any environment-action pair and future rewards. In other words, the role of deep neural networks as *universal function approximators* has been key in recent advances in reinforcement learning.

As far as marketing applications are concerned, although supervised learning applications tend to monopolize the headlights, and true reinforcement learning applications in marketing tend to be scarce, one should not underestimate the promises of reinforcement learning in our field.

The simplest examples of reinforcement learning in marketing are multi-arm bandit problems, where the objective is to "earn while learning," hence optimally balancing exploration and exploitation. Academic literature already contains numerous examples of real-time optimization of website designs (Hauser et al. 2009, 2014, Urban et al. 2014), online advertising (Schwartz et al. 2017, Baardman et al. 2019), and pricing problems (Misra et al. 2018) using this approach.

Compared to the wider field of reinforcement learning in AI, however, these applications could be deemed as relatively simple, since few truly account for *delayed* rewards.

A more compelling application would be a situation where the state of the environment (*S*) is defined as all the information available about a specific customer and the current business

environment, the actions (A) as the marketing actions the firm could take, and the delayed rewards (R) as the *long-term* profitability of said customer. In theory, if we could approximate the function $f(S, A) \rightarrow R$, then all marketing actions (e.g., advertising campaigns, targeting decisions, promotion actions, pricing) could be optimally and truly automated.

Given deep neural networks' ability to tackle complex predictive tasks, this marketing application of reinforcement learning –and many others of the sort– are already within the realm of possibility from a technological point of view. Other reasons currently prevent us from realistically envisaging such an approach yet, though, as we discuss next.

PITFALLS AND DANGERS OF ARTIFICIAL INTELLIGENCE

The major strength of current AI algorithms lies in their ability to uncover hidden patterns in data and to autonomously create higher-degree constructs from raw data, with limited or no human intervention. For instance, a deep learning perceptron can autonomously identify previously-unidentified interactions between predictors, a convolutional network can independently identify and recognize abstract concepts such as "logos" or "eyes" in pictures, and a recurrent neural network can discover important patterns unbeknown to the researcher. That ability has allowed researchers to tackle predictive tasks that were too complex, or involved too many moving parts (i.e., inputs), to be fully specified by a human mind.

AI algorithms do not simply calibrate parameters of a human-specified model; in a sense, they create the model themselves, autonomously. There lies their undeniable strength, which in turn creates a specific set of challenges that we discuss next.

Lack of common sense

A growing field in AI is emotional intelligence (e.g., Mostafa, 2016, Felbermayr and Nanopoulos, 2016), the ability of computers to recognize emotions in humans. Emotional

intelligence applies to image recognition (e.g., detecting happiness in a visage, or finding cues of lies in facial expressions), voice analysis (e.g., detecting an angry customer reaching a call center), or text analysis (e.g., detecting dissatisfaction in an online review, Vo et al., 2018). Given consumer's unwillingness to interact with AI (Longoni et al., 2019; Luo et al., 2019), the ability of AI systems to recognize –and possibly mimic– emotions will likely grow in importance to alleviate consumers' reluctance, too.

There is, however, a conceptually significant difference between recognizing and understanding emotions[4]. A computer program, as evolved as it might be, does not understand joy, let alone feels it. At most, it can be trained to identify and recognize geometric patterns in pictures that are statistically associated with an image category we, humans, arbitrarily labeled "smile." In a sense, AI algorithms are psychopaths; they can be trained to recognize, and even fake emotions, but we are eons away from a time where they may feel them.

The same goes for consciousness or understanding. An AI program may autonomously learn there exist statistical associations between the words "king" and "crown," and be able to use these associations to auto-generate sentences that would seemingly make sense to a human, but would have no understanding of what these words or sentences *truly* mean. Likewise, an autonomous car can be taught to avoid pedestrians at all costs but would do so with no understanding of the intrinsic value of life. To an AI algorithm, hitting a pedestrian is no more than a numerical penalty to avoid –and such penalty, if endured, would generate neither pain, remorse, or guilt.

While AI algorithms lack emotions, consciousness, and understanding, one of the difficulties found in practice is that they also lack common sense –something as obvious to understand as it

[4] In psychology, emotional intelligence distinguishes recognizing, using, understanding, and managing emotions. These abilities are distinct yet related (Colman, 2015).

is easy to forget from time to time–. In other words, they do not obey a set of unsaid rules any human being would implicitly agree upon and have no understanding of the world they operate in. Any unspecified goal or constraint –even the most obvious– is inexistent and will not matter.

If a firm hires a communication agency to design a publicity stunt, bot parties implicitly know that the marketing campaign should not endanger human lives or break the law. In AI, however, there is no such thing as “it goes without saying.”

In traditional statistics, the analyst’s domain knowledge and expertise play an essential role in infusing common sense in the model (e.g., through the selection of relevant independent variables, or the establishment of the causal graph, for instance). This is far less true in AI applications, where data is usually injected in raw form (albeit in large quantities), and the analyst partly outsources higher-degree learning to the model.

This absence of “common sense” makes the specification of objective functions, that we discuss next, a particularly complex and often underrated task.

Objective functions

In reinforcement learning, an objective function specifies the set of rewards (resp. penalties) that the AI algorithm will try to maximize (resp. minimize) over time (Sutton and Barto, 2018). In marketing, researchers and managers commonly set objectives that weigh profit and market share maximization, product cannibalization, customer retention, as well as utility maximization (e.g., Gönül and Hofstede, 2006; Gurca et al., 2002; Natter et al., 2017). Since an AI algorithm is not bounded by common sense, and not limited by a predefined set of features or model specifications to operate from, however, the definition of a holistic objective function becomes of the utmost importance.

As an illustration, one of the authors fondly remembers his first dabbling in the domain of reinforcement learning, some twenty years ago. As a learning exercise, he embarked on the programming of an autonomous spacecraft tasked with exploring a virtual environment.

At first, he set a penalty of -1 every time the artificial agent bumped into an obstacle. After several hours of exploration and reinforcement learning, the agent learned to stay still. Although thought to be a bug at first, it quickly appeared that the algorithm learned the optimal strategy indeed. While devising the exercise, common sense dictated that the agent should move; but since this constraint was not explicitly imposed on the agent, it did without it.

Consequently, the author added a second penalty of -1 every time the agent was remaining still, hence forcing it to explore its environment. After one night of computations and automated trials and errors, the author anxiously checked the results the next morning. To his dismay, the spacecraft was frenetically spinning on itself, which was undoubtedly the optimal strategy. The basic neural network learned to outsmart its creator again.

In a final attempt, the author eventually decided to give a reward to the artificial agent based on the speed it achieved. The next morning, the agent was demonstrating a peculiar behavior: it accelerated as fast as it could in a straight line, without braking or even attempting to turn, violently bumped into an obstacle, stopped, turned around, and repeated the same process. The reward received for speeding was outweighing all the other penalties.

Reflecting on this early experience, a result satisfactory to its creator would have been an agent that *evenly* explored the space and avoided obstacles, while *aesthetically* (however aesthetic may be defined) navigating its environment.

Although simplistic, this example clearly illustrates the difficulties of setting objective functions in environments where some human objectives are implicitly understood, but difficult

to translate into quantitative rewards and penalties. While this challenge applies to supervised learning, this is especially true in the domain of reinforcement learning.

For instance, a badly-programmed –or, more precisely, a badly-*incentivized*– autonomous car instructed by its passenger to "go to the airport ASAP" might reach its destination chased by police cars, with blood on its front hood and vomit on its seats.

At the other extreme, an autonomous car that receives an infinitely-negative penalty for hitting a pedestrian will eventually learn to stay still and not move at all, hence becoming dysfunctional, as in the early experiment described above. In reinforcement learning, companies need to –often explicitly– quantify undesirable outcomes through numerical penalties (the loss of life, a customer churning, breaking the law) in comparison to the rewards for achieving desirable results (arriving on time, maximizing satisfaction, making money). We predict this will lead to quite uncomfortable discussions in the boardroom, and in the public space.

In his canonical thought experiment, Bostrom (2003) develops the case where an artificial general intelligence tasked with the objective of manufacturing as many paperclips as possible, although designed without malice, could ultimately destroy humanity since no reward for humanity's terminal values (such as love, life, and variety) would be built-in in the AI's objective function.

The problem of badly-defined objective functions is not the appanage of AI applications. For instance, Natter et al. (2017) developed a decision-support system for pricing and promotion that aimed at maximizing profit, as initially requested by the sponsor organization. When it became obvious that the model often achieved higher profits at the expense of sales volumes, it led to "emotional discussions with merchandise managers" (p.579). The authors eventually resolved to morph the initial objective function into a weighted function of gross profits, sales, and demand.

The challenge with AI applications is that they can tackle extremely complex and multi-dimensional problems where the analyst does not fully understand causalities; it is both their major strength and their most dangerous weakness. AI models are, therefore, more likely than traditional models to generate unexpected, delayed, and hard-to-quantify consequences, oblivious to the designer's objective function. For instance, in online targeted advertising, Lambrecht and Tucker (2019) demonstrated that since women are more highly-priced target consumers than men, real-time-bidding competition is fiercer. Consequently, women are less likely to be targeted with less-profitable ads, such as announcements for job availabilities in IT, hence creating gender biases in the advertisement of professional opportunities.

In Natter et al. (2017), the managers of the company immediately identified the negative impact on sales volumes and decided to interfere because they understood that prices and volumes were causally linked, and therefore knew what indicators to benchmark. On the other hand, by the time the scientific community determined that COMPAS, an AI-based software used in the judiciary system, was racially biased (Angwin et al., 2016), hundreds of convicts had already suffered the consequences (see the forthcoming section on biased AI).

While reinforcement learning has great potential in the realm of marketing, one must remember that the objectives marketing managers are pursuing are complex, multi-faceted, hard to quantify, and often underthought and badly conceptualized. When a manager considers profit maximization, it is (often) bounded by unspoken assumptions regarding legality, morality, fairness, and ethics. Knowing how effective AI algorithms are at reaching their objectives, marketing managers need to remember that overlooking soft objectives when setting reinforcement learning's rewards and penalties –on the premise that said objectives are hard to quantify– may lead to the business equivalent of "blood on the hood and vomit on the seats."

Safe and realistic learning environment

Our inability to clearly define and quantify a wide range of firm's objectives is not the only impediment to the rapid adoption of reinforcement learning in marketing.

An AI agent will learn to map the function $f(S, A) \rightarrow R$ through thousands, if not millions of trials and errors, balancing exploration and exploitation. The question then becomes "in what kind of environment this learning can take place?"

If a set of predefined and arbitrary rules define the environment, such as in the games of Chess or Go, then the learning environment can be fully simulated to perfection, and an artificial agent can learn by playing against itself. For instance, in its first instance, Alpha Go self-played 1.3M games against itself and self-evaluated 30M positions' winning probabilities (Silver et al., 2016).

If the environment is more complex and realistic, but still adheres to a finite set of well-defined laws, then a computer simulation can still provide an adequate learning environment. For instance, regarding the homing guidance system of an air-to-surface missile, an artificial agent can be trained in a computer simulation that incorporates and models the laws of aerodynamics, the principles of flying (lift, gravity, thrust…), and may even simulate meteorological conditions (e.g., wind).

Such is not the case in marketing. Customers do not adhere to physical laws, and competition does not follow a finite set of well-defined rules. If reinforcement learning can optimally guide the experimentation of random strategies in real life, it is likely to be deemed too costly, too slow, too dangerous, and therefore unacceptable. Consequently, the design of a computer simulation of the environment (i.e., modeling customers' behaviors) remains an essential step in the development and training of effective reinforcement learning approaches.

Underestimating the importance of such a preliminary step may lead to doubtful results. For instance, one of the authors experimented with reinforcement learning to optimize direct marketing decisions; the algorithm learned to solicit some individuals *every day* for months on end. The strategy devised by the autonomous agent was optimal given the simulated environment in which it was being trained, but the environment itself was not realistic enough. In another case, Tkachenko (2015) tentatively reports that the use of a reinforcement learning approach could improve the fundraising of a charity by more than 50%. Knowing that donors in that dataset were already heavily solicited, this result is unlikely to hold in real life, as he recognized.

To create a realistic learning environment in direct marketing, one would need to model customer's fatigue, the unobserved relationship between the number of solicitations and dropout rate, or the likelihood of customers to "sign off" if feeling over-solicited, to name only a few relevant factors. In pricing and promotion applications, one would need to better understand the relationship between price variations, brand perceptions, and ultimately preferences; or between price promotions and stockpiling. One of the authors listened to the complaints of a firms' marketing managers who lamented that "every time we think we know how to do campaigns, the next one fails and a new factor seems to play a role." Unless all these factors are understood (and properly measured), an AI tasked with designing the next campaign is equally doomed to fail.

It is tempting to believe that AI has reduced our needs to understand the causal relationships between inputs and outputs. Given a sufficiently large dataset and a powerful-enough deep learning algorithm, a deep learning model could approximate any causal relationship with neither prior knowledge nor a deep understanding of the phenomenon under scrutiny. Although there is some truth to this statement, AI methods also highlight the pressing needs of the scientific community to more precisely than ever understand the data-generation mechanisms at play and design the scientific tools to manipulate causal inferences (Pearl and Mackenzie, 2018).

In other words, AI applications, especially in reinforcement learning, may promote rather than hinder the need for in-depth consumer behavior theory. So is the case of *biased* and *explainable* artificial intelligence, which we cover next.

Biased artificial intelligence

In a recent case (see Harvard Law Review, 2017), Mr. Loomis was sentenced to six years of imprisonment following a drive-by shooting incident in La Crosse, Wisconsin. In preparation for sentencing, the Wisconsin Department of Corrections produced in court a recidivism risk assessment generated by an AI-driven piece of software called COMPAS.

COMPAS estimates the risk of recidivism based on a criminal reference database and the offender's criminal history. As the methodology behind COMPAS is a trade secret, the Department of Corrections only reported the estimates of recidivism risk to the court.

One of the questions that emerged during the judicial debates was whether or not COMPAS was making racially-prejudiced risk assessments. Some argued that, since individuals' race was *not* keyed into the software to make its predictions, the algorithm could not possibly be biased against black defendants. Angwin et al. (2016) proved otherwise, discovering that "Black defendants who did not recidivate were incorrectly predicted to re-offend at a rate of 44.9%, nearly twice as high as their white counterparts at 23.5% […]. In other words, COMPAS scores appeared to favor white defendants over black defendants by underpredicting recidivism for white and overpredicting recidivism for black defendants."

How is it even technically possible for a piece of software to suffer from prejudices? Or, to cast the discussion in terms of higher-order learning, how is it possible for a deep learning algorithm to autonomously reconstruct the notion of race from the raw data to predict recidivism?

First, it is well established that, in the USA, African-Americans are more likely to be wrongfully convicted (Gross, Possley, and Stephens, 2017). For instance, African-American

prisoners convicted of murder are about 50 percent more likely to be innocent than other convicted murderers, and innocent black people are about 12 times more likely to be convicted of drug crimes than innocent white people. Since the COMPAS predictive model is calibrated on past convictions, it does not predict the risk of recidivism per se, but rather the risk of being *convicted* of recidivism, and there is a consensus in the judicial system that such available, historical data is racially biased.

Second, even though defendants' race is not part of the calibration data (i.e., it is not one of the predictors in the model), other pieces of information can serve as proxies to indirectly inform the algorithm about defendant's race, such as geographical indicators, schools attended, profession, or yearly income.

A sufficiently powerful, AI-driven predictive model would be able to uncover the construct of race in the data autonomously, and hence to make racially-biased predictions. These predictions would consequently be more "accurate" in the sense that they would better fit past historical data. In other words, a powerful AI algorithm shall have no problem reproducing with great accuracy the biases and prejudices found in its training data.

The risk of prejudiced predictions is even true in the context of unsupervised learning. For instance, in word embedding (see the section on unsupervised learning), if the algorithm parses sentences such as "he drove a car" and "he drove a truck," the model will automatically learn from the data that the words "car" and "truck" must share certain features. However, if the word "truck" is more often associated with men than with women, the algorithm will automatically replicate gender prejudices in its predictions (Bolukbasi et al., 2016).

In a marketing context, this could translate into a price-optimization algorithm that learns to charge a higher price for women (Ayres and Siegelmen, 1995), an automatic advertising algorithm that inadvertently targets vulnerable or disadvantaged consumers (UNCTAD, 2018), or

an AI-based salesforce assistant that learns from past data the effectiveness of preying on black youth (Anderson, 2016). These issues have deep legal ramifications since, in the USA, as in many other countries, discrimination does not need to be intentional to be punishable by law.

All ethics considerations set aside, biases in the data might also be the result of endogeneity, rather than prejudice or social biases. In marketing research, endogeneity is a prevalent topic that has received growing academic attention over the years (e.g., Sande and Ghosh, 2018; Rutz and Watson, 2019). Still, even in the field of marketing, its actual impact appears to remain misunderstood by many, and its potential cure misused by most (Rossi, 2014).

We find worrying that a search on the keyword "endogeneity" in computer science, machine learning, and information systems literature leads to an alarmingly low number of hits. This reflects the widespread attitude that AI makes thinking about causal inference obsolete.

Suppose a direct marketing company wishes to deploy an AI-driven algorithm to predict the likelihood of a customer to make a purchase shall she receive a commercial catalog from the company. Endogeneity would creep in the model if, in the past, managers did not target prospective customers randomly. In reality, management probably used some RFM-based rules to select and target those believed to be the most receptive customers at the time (e.g., Elsner et al., 2004).

Such endogeneity in the data –used to calibrate the predictive model– would likely bias the predictions. For instance, it might underestimate the likelihood of an inactive customer to make a purchase, hence reducing the potential profitability of the model. It may even create a downward spiral of self-fulling prophecies, where the model would learn to concentrate on and target fewer and fewer customers over the years, hence optimizing short-term gains along the way, but negatively impacting long-term profits.

The disturbing fact about endogeneity is that, while it is an omnipresent phenomenon that likely plagues numerous predictive endeavors (Rutz and Watson 2019), its presence *enhances* the apparent predictive accuracy of the models. For instance, if the US legal system were indeed racially prejudiced, adding race –or any sociodemographic variables endogenously correlated with race– as an input of the model would make predicting who is going to be sentenced to jail an easier predictive task.

One of the reasons endogeneity appears to receive so little attention in the machine learning community is that many of the predictive tasks researchers are interested in (e.g., predicting the presence of a word in a sentence, or an object in an image) do not entail a decision-feedback loop. For instance, predicting the presence of melanoma on a medical image does not make future melanoma more or less likely.

However, predicting the likelihood that a customer will churn is likely to influence his likelihood of actually churning. If a model predicts a customer will likely leave the company after a service incident, management may use this prediction to decide to cut its losses and redirect marketing investments and loyalty program efforts towards more promising customers, hence increasing even further the chance of that disgruntled customer churning. Future iterations of the AI predictive algorithm will observe (and rightfully predict) an even-higher likelihood of churning for this class of customers, and the self-fulfilling prophecy will come even truer.

The omnipresence of feedback loops in marketing contexts, in the form of data → predictions → decision → data, is sure to create endogeneity issues. It is a well-known phenomenon that is not specific to the use of AI.

AI-based approaches, however, are likely to make this problem more acute, for two reasons. First, AI-based predictive models –specifically deep-learning neural networks– have proven their

superior ability to discover hidden patterns in data, hence making biases and endogeneities even more likely to be picked up by the model. Second, the high complexity of these models transforms them into "black boxes," and many managers have already resolved to accept their predictions at face value. This latter point is a worrying trend which we discuss next.

Explainable artificial intelligence

In a well-known example (Ribeiro et al., 2016), a neural network was tasked to differentiate between pictures of dogs and wolves. While the model accuracy was excellent, it failed to achieve its goal, but instead learned that wolves were often on snow in their picture and dogs were often on grass. It learned to differentiate the two species by looking at the picture environment rather than at the animal features. The model exploited spurious correlations, which was an easier learning path than the one intended by the model builders.

As stated by Ribeiro et al. (2016), "Understanding the reasons behind predictions is, however, quite important in assessing trust, which is fundamental if one plans to take action based on a prediction, or when choosing whether to deploy a new model. Such understanding also provides insights into the model, which can be used to transform an untrustworthy model or prediction into a trustworthy one."

Explainable and interpretable artificial intelligence is a growing field in the AI community, both in size and in strategic importance (e.g., Lipton, 2016; Doshi-Velez and Kim, 2017; Turek, 2017), including in the social sciences (e.g., Miller, 2019). It refers to methods and techniques in AI such as results (e.g., predictions, classifications, recommendations of actions) can be understood by human experts. Specifically, it involves explaining (1) the intention behind the system, (2) the data sources used, and (3) how the inputs are related to the outputs of the model.

The concept of explainable AI is closely linked to the one of "intellectual debt" (Zittrain, 2019). Even if a black-box AI algorithm is accurate and unbiased, not understanding the

theoretical and causal underpinnings of its predictions may cause severe issues down the road, because AI models will increasingly interact with one another, and generate data on their own, hence creating numerous undetected endogeneity issues beyond our control. Our increased ability to generate theory-free predictions may "also involve a shift in the way we think—away from basic science and toward applied technology" (Zittrain, 2019).

Knowing the superior ability of AI systems to leverage endogenous relationships in the data, exploit spurious correlations, replicate human biases, and make theory-free predictions, we argue that management should closely scrutinize AI-based marketing models, and that their transparency, explainability, and interpretability should be a constant and pressing concern[5].

Controllable artificial intelligence

On June 3rd, 2017, a terrorist vehicle-ramming and stabbing took place in London, England. In the panic that ensued, many tried to flee the scene using all modes of transportation available, including subway, cabs, and ride services such as Uber. Following the surge in demand, Uber's pricing algorithm adapted the ride prices automatically, more than doubling the usual fares at that hour, which quickly caused a social uproar. For instance, the Resistance Report (Cahill, 2017) stated that "Ride-sharing company Uber took advantage of the London terrorist attack to make a hefty profit off of people evacuating affected areas."

In truth, the surge in prices was not the result of a decision made by anyone at Uber, but the result of a dynamic pricing algorithm. Despite the social outrage, one could argue that Uber was both well prepared and extremely reactive. First, the company had KPI and monitoring in place to swiftly realize there was a problem. Second, they had preemptively designed mechanisms to override the algorithmic decisions (it took them a few minutes to turn it off). Third, from a PR

[5] We would be remiss not to mention that some authors suggest that, in the case of high-stake decisions, one should avoid using and then trying to explain balack box models altogether, and should rather rely on interpretable models instead from the get go, even at the cost of model accuracy (see Rudin, 2019).

perspective, they quickly communicated about what was going on, made all rides free of charge in the area of the attack, and reimbursed the affected customers within 24 hours.

However, despite their preparedness, Uber’s surge in prices following the London attack has stained the company’s reputation among those affected. This incident should serve as a warning to all that a good AI is a controllable AI. At the very early design stages, marketing organizations *need* to put in place mechanisms to control, stop, and override AI algorithms in real-time (Russell 2019, Armstrong et al. 2012).

In business in general and marketing in particular, there are too many “unknown unknowns” to let AI algorithms, as effective and efficient as they may be in their day-to-day operations, function without human control and overriding capability.

The Paradox of Automation

In her seminal paper, “Ironies of Automation,” Brainbridge (1983) identifies the paradox of automation, which can be summarized as follows: Automation aims at replacing human manual control, planning, and problem-solving by automated processes and devices. After a process has been automated (e.g., through AI), experts are left with responsibilities for two kinds of interventions: verifying and controlling AI operations (the *control* aspect discussed in the previous section), and taking over operations when the situation dictates, such as when unusual circumstances or extreme difficulties arise that AI cannot handle (e.g., Uber in London).

The paradox of automation stipulates that because mundane tasks are the easiest to automate, but performing mundane tasks contributes to hone the human expert’s skills to prepare her to perform more complex ones, automating simple tasks will indeed *deplete* human experts from the knowledge, experience and expertise they need to perform the tasks AI cannot.

The paradox of automation has already made the headlines after the crash of an airplane, highlighting that less-experienced pilots lacked the skills to take over and compensate for faulty

AI (Oliver et al., 2017). In the medical community, the prospect of AI taking over mundane chirurgical operations casts a shadow on surgeons' future abilities to perform more delicate or unique operations, if they do not hone their skills and accumulate experience on routine operations first.

While the paradox of automation will surely imply less dramatic consequences in marketing than in medicine or aerospace, one can neither deny its existence nor underestimate its consequences in our field. One can only wonder what would be the consequences on the quality of the work of customer service agents, sales representatives, content marketing editors, CRM specialists, or target marketing experts if most of the mundane and repetitive tasks they usually hone their skills on (e.g., gaining a deep understanding of customers and their needs) were automated, and they were left dealing exclusively with the extreme cases, difficult problems, and outlier situations?

TACIT KNOWLEDGE TRANSFER AND AI: THE NEXT FRONTIER

Knowledge creation and knowledge transfer

In the first section of this article, we argue that what truly distinguishes recent AI applications (i.e., deep learning) from traditional statistical methods is their ability to **generate higher-order learning**, autonomously, and without relying on human expert knowledge. A deep neural network will discover complex relationships among hundreds of seemingly unrelated indicators to predict the likelihood an online visitor will click on an ad, or discover features in images to predict the likability of a logo, or identify patterns of behaviors in historical purchase data that human experts were oblivious to.

In a sense, the strength of recent AI methods stems from their ability to bypass the need to rely on human experts' knowledge, unlike traditional methods where data scientists transfer their

knowledge of the subject matter to the model through feature extraction, feature engineering, and model specification. When that knowledge is lacking or arduous to articulate, AI shines.

There lies AI’s strength. There also lies AI's major weakness.

Most of the dangers and pitfalls we have discussed in the second section arise from challenges in **knowledge transfer**, either from the expert to the AI model or from the AI model back to the expert:

- The difficulty of articulating basic assumptions and untold constraints, which is common sense to the analyst, but completely foreign to the AI model unless formally specified;
- The struggle of articulating complete and balanced objective functions, including the need to avoid obviously undesirable consequences;
- The difficulty of modeling a realistic, simulated environment in which an AI model could learn an optimal strategy, without experimenting it through trials and errors in real life, and without jeopardizing customer loyalty and satisfaction, or revenues, while doing so;
- The dangers of confounding correlation for causation, or to unwillingly –and unknowingly– reproduce biases present in the data;
- The challenge of understanding “why” a black-box model makes a specific prediction and spotting its errors.
- The paradox of automation, where transferring an increasing number of (possibly mundane) decisions to AI machines may ultimately deplete marketing stakeholders of the knowledge and expertise they need to perform complex tasks.

While AI is likely to have a deep impact on our society at large, anecdotal evidence suggests that it will permeate much of marketing operations as well (Forbes, 2019). In particular, it will have a deep and profound impact on customer experience (Hoyer et al., 2020) and customer

relationship management (Libai et al., 2020). We argue, however, that management, human resources, marketing, and sales are more likely to suffer from AI current limitations than other business fields, such as finance, information systems, operations, or accounting. In particular, marketing deals primarily with customer behavior and human interactions, and as such, relies heavily on tacit knowledge and common sense –a kind of knowledge that is especially challenging to transfer *to* and *from* an AI model–.

Tacit knowledge vs. explicit knowledge

Sternberg et al. (1995) refer to common sense as practical intelligence, i.e., intelligence that is driven by the facile acquisition and use of tacit knowledge. Polanyi (1962) defines tacit knowledge as that which cannot be fully explained even by an expert and can be transferred to another person only through a long period of apprenticeship. Lim (1999) suggests that tacit knowledge comprises of skills and “know-how” that cannot be easily shared. Badaracco (1991) suggests that tacit knowledge exists in individuals or groups of individuals and refers to such knowledge as embedded knowledge.

Formal or explicit knowledge, on the other hand, is captured in manuals and standard operations and then shared with others through either formal training, courses, or books (Lee and Yang, 2000). Edmondson et al. (2003) compare the transfer of complex knowledge (i.e., tacit and context-dependent) to the transfer of simple knowledge (i.e., explicit and context-independent) and found that relatively close relationships and personal contact were important for the former but not for the latter.

Previous studies in marketing have emphasized the importance of tacit knowledge in driving positive marketplace outcomes. For instance, market-oriented firms rely to a large extent on tacit knowledge that arises from greater inter-functional coordination (Madhavan and Grover, 1998)

through formal and informal processes that help the dissemination of intelligence across the firm (Jaworski and Kohli, 1993). In particular, knowledge transfer that occurs through informal contacts is critical for firms to respond effectively, especially in environments characterized by high environmental uncertainty (Gupta and Govindarajan, 1991). Employees' tacit knowledge is also considered a major competitive advantage in relationship marketing (Pereira et al., 2012).

Transfer of tacit knowledge

Tacit knowledge plays an important role in marketing, and its flow within the organization and across marketing functions is a key driver of a firm's competitiveness. The importance of tacit knowledge transfer within a marketing or sales organization (e.g., Atefi et al., 2018, Chan et al., 2014) applies to AI modeling efforts as well. Consequently, marketing organizations should diligently focus on the transfer of tacit knowledge from various marketing stakeholders to the AI algorithm (to design proper objective functions, identify untold assumptions and objectives, alleviate endogeneity); but also from the AI algorithm back to the experts (to contribute to institutional learning, and ensure explainability and controllability).

Nonaka's (1994) four-step process is particularly useful as it explains how new knowledge is created and managed through the repeated transfer of tacit knowledge to explicit knowledge and back. According to this framework, the transfer of information embedded in emotions and nuanced contexts heavily relies on the transfer of tacit knowledge through observation and experience, not language and formal concepts. Accordingly, AI applications must develop the capability to gather the information that is not necessarily written or archived, and incorporate this new information in existing rules and applications.

Chess experts have long reached that conclusion. After the phenomenal victories of Alpha Zero, chess coach Peter Nielsen was reported to say that "the aliens came and *showed* us how to play chess" (italics added). While Google chess program learned to play in innovative ways that

experts never considered viable before, human beings have, in turn, learned from it and improved their games through repeated interactions. In other words, tacit knowledge transfer could be achieved "not through language, but by observation, imitation, and practice" (Nonaka, 1994, p.19).

The same process could –and should– be applied to marketing. For instance, AI algorithms embedded in intelligent agents that interact with consumers could capture tacit knowledge, specifically the underlying motivations and goals that drive individuals' behavior, through repeated interactions. Another use case from a firm perspective could be the use of AI-driven marketing technologies that capture tacit knowledge possessed by the sales team to improve the effectiveness of marketing efforts.

As AI-driven machines capture tacit knowledge, this information can be first tested and then employed in AI algorithms that balance the objectives of various stakeholders in the ecosystem. Such an approach will help AI applications identify relevant causal factors, alleviate concerns on endogeneity, and thus build new explicit or formal knowledge. The cycle of knowledge creation continues through the internalization or learning phase of Nonaka's framework, where newly created explicit knowledge can help generate new tacit knowledge and/or update existing tacit knowledge.

In his seminal paper, Nonaka (1994) wrote that "At a fundamental level, knowledge is created by individuals." Recent advances in artificial intelligence, and more specifically in deep neural networks, challenge this statement.

AI ability to generate higher-order learning from raw data and self-generated experience, without relying on human expertise, has opened possibilities that were deemed unthinkable less than a decade ago. Today, AI beats human experts at diagnosing skin cancer, recognizing objects,

playing poker, go and chess, and identifying issues in legal documents; and it is already better than average human beings at reading text, driving cars, and writing pop songs.

But AI's ability to operate impermeably from experts' tacit knowledge may also be its biggest weakness in domains where tacit knowledge plays a critical role. This is the case for a wide range of marketing domains such as advertising, brand management, customer experience, engagement and loyalty, international marketing and joint ventures, service quality, and brand portfolio management, to name a few.

In other words, the ability of AI to generate knowledge from the confinement of a black box is remarkable (cf. section 1); but a black box that remains impervious to knowledge transfer will continue to pose significant threats and challenges (cf. section 2). Additionally, we believe these challenges will plague marketing, management, and sales to a greater extent than they will plague other business functions such as finance, operations or accounting, because of the ubiquitous role of tacit knowledge in said domains.

To transfer tacit knowledge from human beings (marketing experts, front-desk employees, sales representatives, and consumers alike) to machines will be paramount to create new products, services, solutions, and relationships. In turn, to transfer what the machine has "learned" back to the marketing experts will be critical to identify biases and errors, encourage humans to place greater trust in AI machines, and accept their decisions with greater conviction. Or, on the contrary, thanks to a deep understanding of how AI operates (understanding obtained through regular observations, interactions, and practice), experts might be able to spot when AI machines go astray in face of unusual circumstances (e.g., think how faster Uber might have reacted in the wake the London terrorist attack if "AI intimacy" was engrained in the organization).

If AI machines fail to incorporate tacit knowledge in their algorithms, and if they fail to transfer what they have learned back to human beings efficiently, there would be greater confusion and inefficiencies, thus defeating the very purpose of an intelligent machine.

CONCLUSIONS AND RESEARCH AGENDA

The marketing literature teaches us that an organization can only achieve successful tacit knowledge transfer through shared experience and proximity. Consequently, marketing organizations should seek to facilitate and systematize interactions between AI and marketing stakeholders, and create an ecosystem to foster a form of "intimacy" between AI and experts (or consumers) through two-way observation, imitation, and practice. Tacit knowledge transfer should occur both ways, because as Michael Polanyi (1996, p.4) put it, even we, human beings, "can know more than we can tell."

For instance, it is symptomatic to note that, in their thorough review of the literature on service robots in the frontline, not once do Wirtz et al. (2018) identify the possibility of robots to learn from frontline employees or vice versa. The literature seems to assume that service robots can only learn from an organization-wide knowledge base and from trials and errors. We argue that to do without frontline employees accumulated tacit knowledge may very well lead to suboptimal performance, and even failures. The industry tendency to equip frontline employees with AI (rather than replace them by AI) seems to have acknowledge that limitation (Grewal et al., 2020).

In terms of research agenda, this article highlights two research areas that seem to be currently underinvestigated –although not completely overlooked– in AI. First, how can humans more effectively transfer their tacit knowledge into AI machines? If AI could rely more heavily on human expertise and understanding and did not need to learn everything from scratch

anymore, such knowledge transfer might alleviate the immense need for data, and help infer causalities in the data (rather than mere correlations), addressing many of the pitfalls discussed in this article. Such knowledge transfer would also be essential to design complex, complete, and accurate managerial objective functions.

Second, the ability to transfer knowledge constructs that have been autonomously generated by AI algorithms back to humans would be equally important to build trust, enhance control, and create a positive feedback loop at the organization level. Such knowledge transfer may rely on data visualisation and managerial training, but would probably also require a form of "AI intimacy" that remains to be developed.

Every time AI has failed to live up to its promises, the field has experienced twice what many call an "AI winter" (Haenlein and Kaplan, 2019), where AI initiatives go through a more inactive cycle, and investments in R&D decrease. As researchers and firms apply AI technologies to an ever-expanding list of business endeavors, we predict that our current inability to effectively transfer tacit knowledge back and forth between AI machines and marketing stakeholders will become increasingly apparent. Unless we work diligently on bridging that gap, and stop designing AI machines as formidable but impermeable black boxes, that realization may sound the death knell of AI in marketing domains where tacit knowledge plays a central role –a third AI winter of sorts–.

REFERENCES

Aggarwal, Charu C. (2018), Neural Networks and Deep Learning: A Textbook, Springer.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb (2019), Prediction Machines: the simple economics of artificial intelligence, Harvard Business Review Press.

Anderson, Lorin W., and David R. Krathwohl (2001), A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives, Allyn and Bacon.

Anderson, Melinda D. (2016), "When For-Profit Colleges Prey on Unsuspecting Students," The Atlantic, published October 24, 2016.

Angwin, J., J. Larson, S. Mattu and L. Kirchner (2016), "Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks," ProPublica, 23 May 2016; available at www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Ansari, A., & Riasi, A. (2016). Modelling and evaluating customer loyalty using neural networks: Evidence from startup insurance companies. *Future Business Journal*, *2*(1), 15-30.

Armstrong, Stuart, Anders Sandberg, and Nick Borstrom (2012), "Thinking Inside the Box: Controlling and Using an Oracle AI," *Minds and Machines*, 22(4).

Atefi, Yashar, Michael Ahearne, James G Maxham III, D. Todd Donavan, and Brad D. Carlson (2018), "Does Selective Sales Force Training Work?," *Journal of Marketing Research*, 55(5), pp.722-737.

Ayres, Ian, and Peter Siegelman (1995), "Race and Gender Discrimination in Bargaining for a New Car," American Economic Review, 85(3), 304-321.

Baardman L., E. Fata, A. Pani, G. Perakis (2019), "Learning Optimal Online Advertising Portfolios with Periodic Budgets," SSRN Scholarly Paper ID 3346642, Social Science Research Network, Rochester, NY.

Badaracco, J. (1991), "Alliances speed knowledge transfer," *Planning Review*, 19(2), pp.10-16.

Bini, S. A. (2018). Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care?. *The Journal of arthroplasty*, *33*(8), 2358-2361.

Bloom, B. S., M. D. Engelhart, E. J. Furst, W. H. Hill, and D. R. Krathwohl (1956), Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: David McKay Company.

Bolukbasi Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai (2016), "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word

Embeddings," in Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

Bostrom, Nick (2003). "Ethical Issues in Advanced Artificial Intelligence," Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence, Vol. 2, ed. I. Smit et al., Int. Institute of Advanced Studies in Systems Research and Cybernetics, 2003, pp. 12-17.

Brainbridge, Lisanne (1983), "Ironies of Automation," *Automatica*, 19(6), pp.775-779.

Brooks, Rodney Allen (1991), "Intelligence without Reason," in Proceedings of the 12th International Joint Conference on Artificial Intelligence, ed. M. Ray and J. Reiter (Sydney, Australia: Morgan Kaufmann, 1991), pp. 569-595.

Cahill, Tom (2017), "What Uber did after the London attack is making people furious," Resistance Report, published June 4, 2017, available at http://archive.is/pBsJE.

Chan, Tat, Jia Li, and Lamar Pierce (2014), "Learning from Peers: Knowledge Transfer and Sales Force Productivity Growth," *Marketing Science*, 33(4), pp.463-484.

Chapmann, Joshua (2017), Neural Networks: Introduction to Artificial Neurons, Backpropagation Algorithms and Multilayer Feedforward Networks, Advanced Data Analytics, Volume 2.

Cho, Kyunghyun, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio (2014). "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation". arXiv:1406.1078.

Colman, Andrew M. (2015), A Dictionary of Psychology, 4th Edition, Oxford University Press.

Da Silva, I. N., Spatti, D. H., Flauzino, R. A., Liboni, L. H. B., & dos Reis Alves, S. F. (2017). Artificial neural networks. *Cham: Springer International Publishing*.

Di Persio, L., & Honchar, O. (2016). Artificial neural networks architectures for stock price prediction: Comparisons and applications. *International journal of circuits, systems and signal processing*, *10*, 403-413.

Dimitrieska, S., Stankovska, A., & Efremova, T. (2018). Artificial Intelligence and Marketing. *Entrepreneurship*, *6*(2), 298-304.

Doshi-Velez, Finale, and Been Kim (2017), "Towards A Rigorous Science of Interpretable Machine Learning," arXiv:1702.08608.

Edmondson, Amy C., Ann B. Winslow, Richard M. J. Bohmer, and Gary P. Pisano (2003) "Learning How and Learning What: Effects of Tacit and Codified Knowledge on Performance Improvement Following Technology Adoption," *Decision Sciences*, 34(2), pp.197-224.

Egriogglu, E., Aladag, C., & Gunay, S. (2008). 'A new model selection strategy in artificial neural networks', *Applied Mathematics and Computation* (195) 591-597.

Elsner, Ralf, Manfred Krafft, and Arnd HuchZermeier (2004), "Optimizing Rhenania's Direct Marketing Business Through Dynamic Multilevel Modeling (DMLM) in a Multicatalog-Brand Environment," *Marketing Science*, 23(2), 192-206.

Erevelles Sunil, Nobuyuki Fukawa, and Linda Swayne (2016), "Big Data consumer analytics and the transformation of marketing," *Journal of Business Research*, 69(2), pp.897-904.

Felbermayr, Armin, and Alexandros Nanopoulos (2016), "The Role of Emotions for the Perceived Usefulness in Online Customer Reviews," Journal of Interactive Marketing, 36, 60-76.

Fine, Terrence (1999), Feedforward Neural Network Methodology, Springer, Information Science and Statistics series.

Forbes (2019), "How Artificial Intelligence is Transforming Digital Marketing," found at *https://www.forbes.com/sites/forbesagencycouncil/2019/08/21/how-artificial-intelligence-is-transforming-digital-marketing*

Gers, F.A., J. Schmidhuber, and F. Cummins (1999), "Learning to Forget: Continual Prediction with LSTM," Published in: 1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)

Goertzel, B. (2015). Artificial General Intelligence: Concept, State of the Art, and Future Prospects. *Journal of Artificial General Intelligence, 5(1),* 1-46.

Gönül, Füsun F., Frenkel Ter Hofstede (2006), "How to Compute Optimal Catalog Mailing Decisions," *Marketng Science*, 25(1), 65-74.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016), Deep Learning, Adaptive Computation and Machine Learning series, The MIT Press.

Grewal, Dhruv, Mirja Kroschke, Martin Mende, Anne L. Roggeveen, and Maura L. Scott (2020), "Frontline Cyborgs at Your Service: How Human Enhancement Technologies Affect Customer Experiences in Retail, Sales, and Service Settings," *Journal of Interactive Marketing*, forthcoming.

Gross, Samuel R., Maurice Possey and Klara Stephens (2017), "Race and Wrongful Convictions in the United States." National Registry of Exonerations, Newkirk Center for Science and Society, University of California Irvine, Irvine, California 92697. March 7, 2017.

Gupta, Anil K., and Vijay Govindarajan (1991), "Knowledge Flows and the Structure of Control within Multinational Corporations," *The Academy of Management Review*, 16(4), pp.768-792

Harvard Law Review (2017), "Criminal Law – Sentencing Guidelines – Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing," March 2017, 130(5), pp.1530-1537.

Hauser J. R., G. G. Liberali, G. L. Urban (2014), "Website Morphing 2.0: Switching Costs, Partial Exposure, Random Exit, and When to Morph," Management Science, 60(6).

Hauser J. R., G. L. Urban, Guy Liberali, M. Braun (2009), "Website Morphing," Marketing Science 28(2).

Haenlein, Michael, and Andreas Kaplan (2019), "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence," *California Management Review*, 61(4), pp.5-14.

Hoyer, Wayne D., Mirja Kroschke, Bernd Schmitt, Karsten Kraume, and Venkatesh Shankar (2020), "Transforming the Customer Experience through New Technologies," *Journal of Interactive Marketing*, forthcoming.

Ismail, Mohammad, Mohd Awang, Mohd Rahman, and Mokhairi Makhtar (2015), "A Multi-Layer Perceptron Approach for Customer Churn Prediction," International Journal of Multimedia and Ubiquitous Engineering, 10(7), pp.213-222.

Jaworski, Bernard J., and Ajay K. Kohli (1993), "Market Orientation: Antecedents and Consequences," *Journal of Marketing*, 57(3), pp.53-70.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260.

Kaplan, Andreas and Michael Haenlein (2019), "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence," Business Horizons, 62(1), Pages 15-25.

Kelleher, John D. (2019), Deep Learning, MIT Press Essential Knowledge series, The MIT Press.

Khan, Salmna, Hossein Rahmani, and Syed Afaq Ali Shah (2018), "A Guide to Convolutional Neural Networks for Computer Vision," Morgan & Claypool Publishers.

Kumar et al. (2020), "Digital Analytics: Modeling for Insights and New Methods," *Journal of Interactive Marketing*, forthcoming.

Kurzweil, Ray (2005), The Singularity Is Near: When Humans Transcend Biology, The Viking Press, 1st edition.

Lambrecht, Anja, and Catherine Tucker (2019), "Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads," Management Science, 65(7).

Lee, Ching Chyi, and Jie Yang (2000), "Knowledge Value Chain," *Journal of Management Development*, 19(9), pp.783-794

Legg, Shane and Marcus Hutter (2007), "A Collection of Definitions of Intelligence," In *Proceedings of the 2007 conference on Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms: Proceedings of the AGI Workshop 2006*, Ben Goertzel and Pei Wang (Eds.). IOS Press, Amsterdam, The Netherlands, 17-24.

Libai, Barak, Yakov Bart, Sonja Gensler, Charles Hofacker, Andreas Kaplan, Kim Kötterheinrich, and Eike Benjamin Kroll (2020), "Brave New World? On AI and the Management of Customer Relationships," *Journal of Interactive Marketing*, forthcoming.

Lim, K.K. (1999), "Managing for quality through knowledge management," *Total Quality Management*, 10(415), pp.615-22.

Lipton, Zachary C. (2016), "The Mythos of Model Interpretability," presented at 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY.

Longoni, Chiara, Andrea Bonezzi, and Carey K. Morewedge (2019), "Resistance to Medical Artificial Intelligence," *Journal of Consumer Research*, 46(4), 629–650.

Luo, X., S. Tong, Z. Fang, Z. Qu (2019), "Machines versus Humans: The Impact of AI Chatbot Disclosure on Customer Purchases," *Marketing Science*, Forthcoming.

Madhavan, Ravindranath, and Rajiv Grover (1998), "From Embedded Knowledge to Embodied Knowledge: New Product Development as Knowledge Management," Journal of Marketing, 62(4), pp.1-12.

Mandic, Danilo, and Jonathon Chambers (2020), Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability, Wiley.

Miller, Tim (2019), "Explanation in Artificial Intelligence: Insights from the Social Sciences," Artificial Intelligence, 267 (February), pp.1-38.

Misra K, E. M. Schwartz, J. Abernethy (2018), "Dynamic Online Pricing with Incomplete Information Using Multi-Armed Bandit Experiments," SSRN Scholarly Paper ID 2981814, Social Science Research Network, Rochester, NY.

Mostafa M., Crick T., Calderon A.C., Oatley G. (2016), "*Incorporating Emotion and Personality-Based Analysis in User-Centered Modelling*," In: Bramer M., Petridis M. (eds) Research and Development in Intelligent Systems XXXIII. SGAI 2016. Springer, Cham

Murray, G., & Wardley, M. (2014). The math of modern marketing: How predictive analytics makes marketing more effective. IDC White Paper. Retrieved from http://www.sap.com/bin/sapcom/en_ us/downloadasset.2014-06-jun-12-15.the-math-of-modern-marketing-how-predictive-analyticsmakes-marketing-more-effective-pdf.bypassReg.html

Natter Martin, Thomas Reutterer, Andreas Mild, and Alfred Taudes (2007), "Practice Prize Report: An Assortmentwide Decision-Support Systems for Dynamic Pricing and Promotion Planning in DIY Retailing," *Marketing Science*, 26(4), 576-583.

Nonaka, Ikujiro (1994), "A Dynamic Theory of Organizational Knowledge Creation," *Organization Science*, 5(1), pp.14-37.

Oliver, Nick, Thomas Calvard, and Kristina Potočnik (2017), "The Tragic Crash of Flight AF447 Shows the Unlikely but Catastrophic Consequences of Automation," Harvard Business Review, September 15, 2017.

Pavaloiu, A. (2016). The Impact of Artificial Intelligence on Global Trends. Journal of Multidisciplinary Developments, 1(1), 21-37.

Pearl, Judea, and Dana Mackenzie (2018), The Book of Why: The New Science of Cause and Effect, Basic Books Eds.

Pereira, Carla, João Ferreira, and Helena Alves (2012), "Tacit Knowledge as Competitive Advantage in Relationship Marketing: A Literature Review and Theoretical Implications," Journal of Relationship Marketing, 11(3), pp.172-197.

Polanyi, M. (1962), Personal Knowledge – Towards a Post-Critical Philosophy, University of Chicago press, Chicago.

Polanyi, M. (1966), The Tacit Dimension, London: Routledge & Kegan Paul.

Power, D. J. (2016). "'Big Brother' Can Watch Us," *Journal of Decision Systems*, 25(1), 578-588.

Ramchoun, H., Idrissi, M. A. J., Ghanou, Y., & Ettaouil, M. (2016). Multilayer Perceptron: Architecture Optimization and Training. IJIMAI, 4(1), 26-30.

Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin (2016), ""Why Should I Trust You?": Explaining the Predictions of Any Classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.1135-1144.

Riikkinen, M., Saarijärvi, H., Sarlin, P., & Lähteenmäki, I. (2018). Using artificial intelligence to create value in insurance. *International Journal of Bank Marketing*, *36*(6), 1145-1168.

Rossi, Peter E. (2014), "Even the Rich Can Make Themselves Poor: A Critical Examination of IV Methods in Marketing Applications," Marketing Science, 33(5), pp.621-762.

Rudin, Cynthia (2019), "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," *Nature Machine Intelligence*, Vol 1, 206-215.

Russell, S. J., and Norvig, P. (2016), Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited.

Russell, S. J. (2019), Human compatible: artificial intelligence and the problem of control, Viking, NY.

Rutz, Oliver J., and George F. Watson IV (2019) "Endogeneity and marketing strategy research: an overview," Journal of the Academy of Marketing Science, 47(3), pp.479-498.

Sande, Jon Bingen, and Mrinal Ghosh (2018), "Endogeneity in Survey Research," International Journal of Research in Marketing. 35(2), pp185-204.

Sarkar, Mainak, and Arnaud De Bruyn (2019), "Predicting Customer Behavior with LSTM Neural Networks," Marketing Science Conference 2019, INFORMS, Rome, Italy.

Schwartz E. M., E. T. Bradlow, P. S. Fader (2017), "Customer Acquisition via Display Advertising Using Multi-Armed Bandit Experiments," Marketing Science 36(4).

Shieber, S. M. (Ed.). (2004). The turing test: Verbal behavior as the hallmark of intelligence. Cambridge: MIT Press.

Silver D., Huang A., Maddison CJ., Guez A., Sifre L., van den Driessche G., Schrittwieser J., Antonoglou I., Panneershelvam V., Lanctot M., Dieleman S., Grewe D., Nham J2, Kalchbrenner N., Sutskever I2, Lillicrap T., Leach M., Kavukcuoglu K., Graepel T., Hassabis D. (2016), "Mastering the game of Go with deep neural networks and tree search," Nature, 529(7587), pp.484-489.

Sternberg, R. J., Wagner, R. K., Williams, W. M., & Horvath, J. A. (1995), "Testing common sense," *American Psychologist*, 50, pp.912-927.

Sutton, Richard S., and Andrew G. Barto (2018), Reinforcement Learning: An Introduction, Second Edition, MIT Press, Cambridge, MA.

Thórisson, K. R., Bieger, J., Schiffel, S., & Garrett, D. (2015, July). Towards flexible task environments for comprehensive evaluation of artificial intelligent systems and automatic learners. In *International Conference on Artificial General Intelligence* (pp. 187-196). Springer, Cham.

Tkachenko, Yegor (2015), "Autonomous CRM Control via CLV Approximation with Deep Reinforcement Learning in Discrete and Continuous Action Space," technical report available at https://arxiv.org/abs/1504.01840.

Turek, Matt (2017), "Explainable Artificial Intelligence (XAI)," DARPA (Defense Advanced Research Projects Agency), available at https://www.darpa.mil/program/explainable-artificial-intelligence.

UNCTAD Secretariat (2018), "Working Group on Vulnerable and Disadvantaged Consumers," Intergovernmental Group of Experts on Consumer Law and Policy (IGE Consumer).

Urban G. L., G. G. Liberali, E. MacDonald, R. Bordley, J. R. Hauser (2014), "Morphing Banner Advertising," Marketing Science 33(1).

Valentin, Jan, and Thomas Reutterer (2019), "From RFM to LSTM: Predicting Customer Future with Autoregressive Neural Networks," Marketing Science Conference 2019, INFORMS, Rome, Italy.

Vo, A. D., Nguyen, Q. P., & Ock, C. Y. (2018). Opinion–aspect relations in cognizing customer feelings via reviews. *IEEE Access*, *6*, 5415-5426.

Wang, K., Zhang, D., Li, Y., Zhang, R., & Lin, L. (2016). Cost-effective active learning for deep image classification. *IEEE Transactions on Circuits and Systems for Video Technology*, *27*(12), 2591-2600.

Wirtz, Jochen, Paul G. Patterson, Werner H. Kunz, Thorsten Gruber, Vinh Nhat Lu, Stefanie Paluch, and Antje Martins (2018), "Brave new world: service robots in the frontline," *Journal of Service Management*, 29(5), pp.907-931.

Zhao, Z., Xu, S., Kang, B. H., Kabir, M. M. J., Liu, Y., & Wasinger, R. (2015). Investigation and improvement of multi-layer perceptron neural networks for credit scoring. *Expert Systems with Applications*, *42*(7), 3508-3516.

Zheng, Alice, and Amanda Casari (2018), Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists, O'Reilly Media.

Zittrain, Jonathan (2019), "The Hidden Costs of Automated Thinking," The New Yorker, July 23, 2019, available at https://www.newyorker.com/tech/annals-of-technology/the-hidden-costs-of-automated-thinking.